Military Speech Communications over Vocoders in Tandem

Mark A. Ericson, Brian D. Simpson, and Richard L. McKinley
Air Force Research Laboratory (AFRL/HECB)
2610 Seventh Street, Bldg 441
Wright-Patterson, AFB, OH 45433-7901

mark.ericson@wpafb.af.mil

VOCODERS IN TANDEM

Speech intelligibility of two types of vocoders was measured using the modified rhyme test. One type of vocoder, a continuous variable slope delta (CVSD), was a waveform encoder. The other type, an advanced multi-band excitation (AMBE), was a parametric encoder. In the first experiment, clear speech was processed through the vocoders. Intelligibility was measured in a control condition, i.e. without vocoding, with each type alone and with two vocoders in tandem. AMBE and CVSD performed similarly, 92.6 and 90.4%, respectively. CVSD-to-AMBE had little effect on intelligibility, measured at 89.2%. However, AMBE-to-CVSD had a large degrading effect on intelligibility. The AMBE-to-CVSD direction scored about 81.7% intelligibility with clear, unaltered speech signals. The asymmetry between waveform-to-parametric and parametric-to-waveform encoders underscores the non-linear nature of tandem vocoders on intelligibility. When vocoders of the same type were in tandem, there was no additional effect on intelligibility. The double CVSD condition yielded 92.2% intelligibility and the double AMBE condition yielded 91%. The deleterious effects of speech clipping were measured in a second experiment, as these are ubiquitous in military radio transmission systems. The AMBE parametric vocoder performed at the 88% level in isolation and at 84% when tandemed with the CVSD waveform vocoder. Alternative methods of encoding speech signals are being explored to improve speech intelligibility performance in military communication systems.

1.0 INTRODUCTION

There are two general classes of vocoders used in military communication systems today; these are parametric and waveform vocoders. Waveform vocoding techniques, such as continuous variable slope delta (CVSD), are highly resistant to noise and bit error effects [1, 2]. Parametric encoders, such as advanced multi-band excitation (AMBE) [3], greatly reduce signal bandwidth, which is helpful in reducing encryption processing requirements and the cost of transmitting a wide band signal. Both waveform and parametric vocoders can provide good speech intelligibility alone at adequate bandwidths [4]. However, a “staging” or “tandem” problem occurs when waveform encoders and parametric encoders are placed in sequence in a given communication system [5]. The distortion of the speech waveform produced by the first vocoder causes the second vocoder to severely distort the speech waveform, thereby reducing the overall intelligibility.

Vocoder algorithms have typically been developed to reduce bandwidth for long distance or secure communications [6, 7]. These devices are not necessarily designed to be compatible in conjunction with other types of vocoders. Military communication systems are likely to have legacy equipment, which will include parametric [7] and waveform types of vocoders [8]. In future military operations, speech communications are likely to occur among more operators from multiple points in the chain of command.

Ericson, M.A.; Simpson, B.D.; McKinley, R.L. (2005) Military Speech Communications over Vocoders in Tandem. In *New Directions for Improving Audio Effectiveness* (pp. 11-1 – 11-8). Meeting Proceedings RTO-MP-HFM-123, Paper 11. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.aps.

Report Documentation Page					*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE **01 APR 2005**		2. REPORT TYPE **N/A**			3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Military Speech Communications over Vocoders in Tandem**					5a. CONTRACT NUMBER
					5b. GRANT NUMBER
					5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)					5d. PROJECT NUMBER
					5e. TASK NUMBER
					5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory (AFRL/HECB) 2610 Seventh Street, Bldg 441 Wright-Patterson, AFB, OH 45433-7901**					8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					10. SPONSOR/MONITOR’S ACRONYM(S)
					11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**					
13. SUPPLEMENTARY NOTES **See also ADM001856, New Directions for Improving Audio Effectiveness (Nouvelles orientations pour l’amelioration des techniques audio)., The original document contains color images.**					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The tandem problem will potentially increase as remote-controlled air vehicles become more numerous in military operations. Operators of remote-control reconnaissance and attack air vehicles, such as the Global Hawk and uninhabited combat aerial vehicles (UCAVS), must communicate with civilian air traffic controllers and military command and control personnel. Ground troops equipped with satellite phones will need to communicate with other military operators via multiple communication links. Achieving good speech intelligibility over such multiple-link communication systems will be critical for safely and efficiently accomplishing military missions.

The test objective was to measure the effects of continuous variable slope delta (CVSD) and advanced multi-band excitation (AMBE) vocoding algorithms on speech intelligibility. These components are considered the critical links in the air traffic controller (ATC) to UAV ground control station communication path. A typical communication path is depicted in Figure 1. Note that the communication can go in both directions from the air traffic controller and the UAV ground control station. The direction of the path determines the order of the vocoders in tandem.

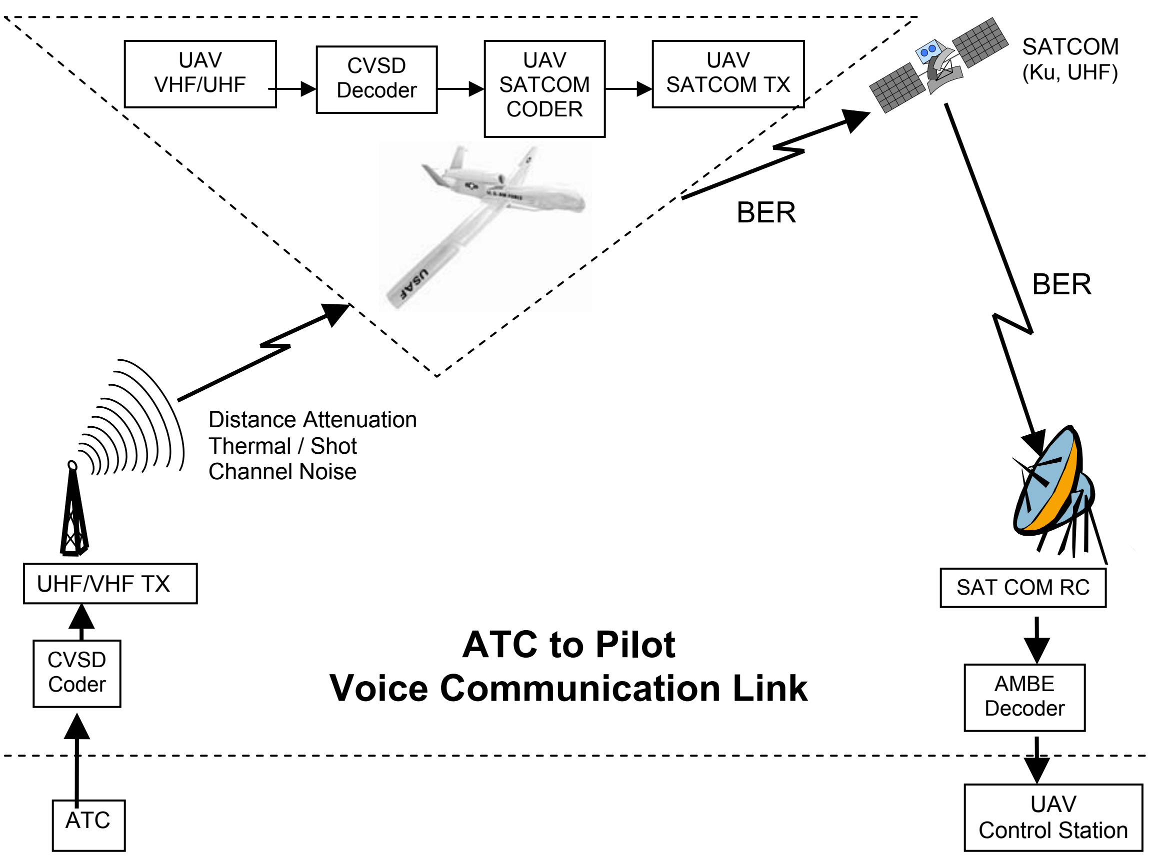

Figure 1: Depiction of communication paths from air traffic controllers to a UAV command and control station

2.0 METHODS

2.1 Equipment

The Air Force Research Laboratory's Battlespace Acoustics Branch (AFRL/HECB) operates and maintains unique facilities for researching and developing voice communication systems for military operating environments. The Voice Communication Research and Evaluation System (VOCRES) [9] is capable of simulating all parts of the military communication path from the talker, via the medium, to the receiver. The six modified rhyme test (MRT) [10] lists of 50 words each were read by three male and three female talkers and recorded in 16 bit format. The speech stimuli were pre-processed off-line using a Windows 98 PC, a 4.8 kbps AMBE processor, and a 16 kbps CVSD algorithm. The processed speech files were played back to the panel of professional listeners. The two types of vocoders were staged together and in both directions of the communication path. Listeners wore H-157A headsets at the response desks of VOCRES.

2.2 Subjects

Five listeners participated in the vocoder studies. The paid volunteer subjects ranged in age from 18 to 51 years with a mean age of 27 years. Each volunteer subject had normal hearing threshold levels and consented to participate in the speech intelligibility experiments.

2.3 Procedures

The degrading effects of the vocoding by single and tandem systems on speech intelligibility were measured using the MRT. The MRT is one of three standardized procedures for measuring the intelligibility of speech over communication systems. The speech utterances were recorded with the test word imbedded in a carrier phrase to reduce the deleterious effects of the attack portion of the automatic gain control circuitry on the intelligibility of the initial consonant. In each session, six pre-recorded talkers processed with the vocoders were played to listeners in a quiet environment. Responses were automatically collected and scored by the computers in the VOCRES facility. Baseline conditions were tested in which each talker's voice was processed with each class of vocoder in isolation and in tandem with each other.

3.0 RESULTS

3.1 Unaltered Speech Vocoded in Tandem

The first experiment determined the performance of the vocoders with unaltered (clear) speech signals. Raw data scores were corrected for guessing by using the formula of 2.4 times the raw number correct out of 50 minus 20. The mean and standard deviations for percent correct intelligibility are plotted in Figure 2. The percent correct values of speech intelligibility were measured to be 96.3% for the control condition of no vocoders, 92.6% for AMBE-alone, 90.4% for CVSD-alone, 89.2% for CVSD-AMBE, and 81.7% for AMBE-CVSD. In the like-vocoder condition, speech intelligibility was found to be 92.2% for CVSD-to-CVSD and 92.6% for AMBE-to-AMBE.

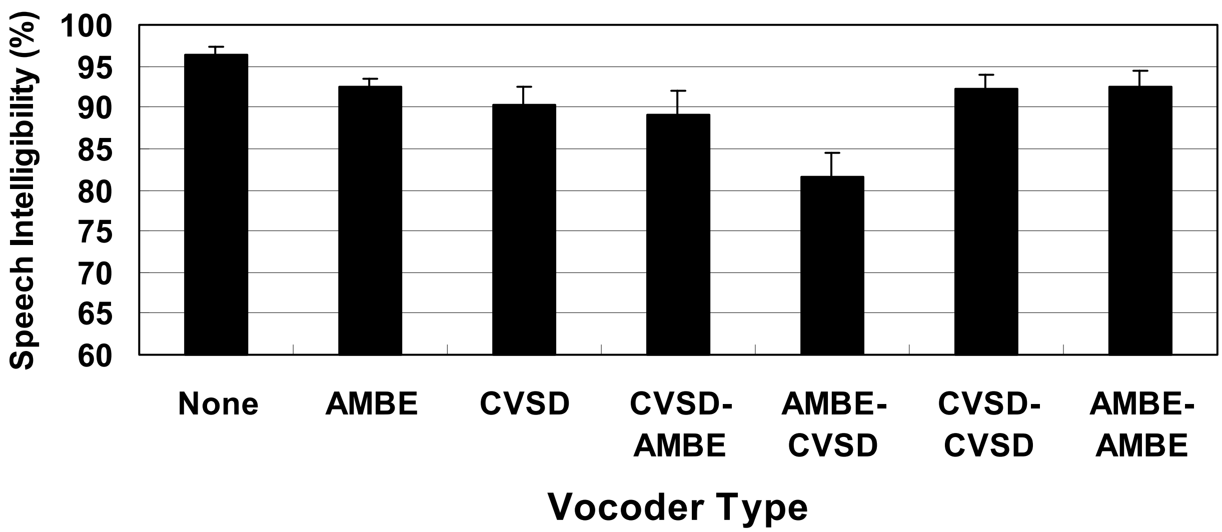

Figure 2: Speech intelligibility versus vocoder type with unaltered speech

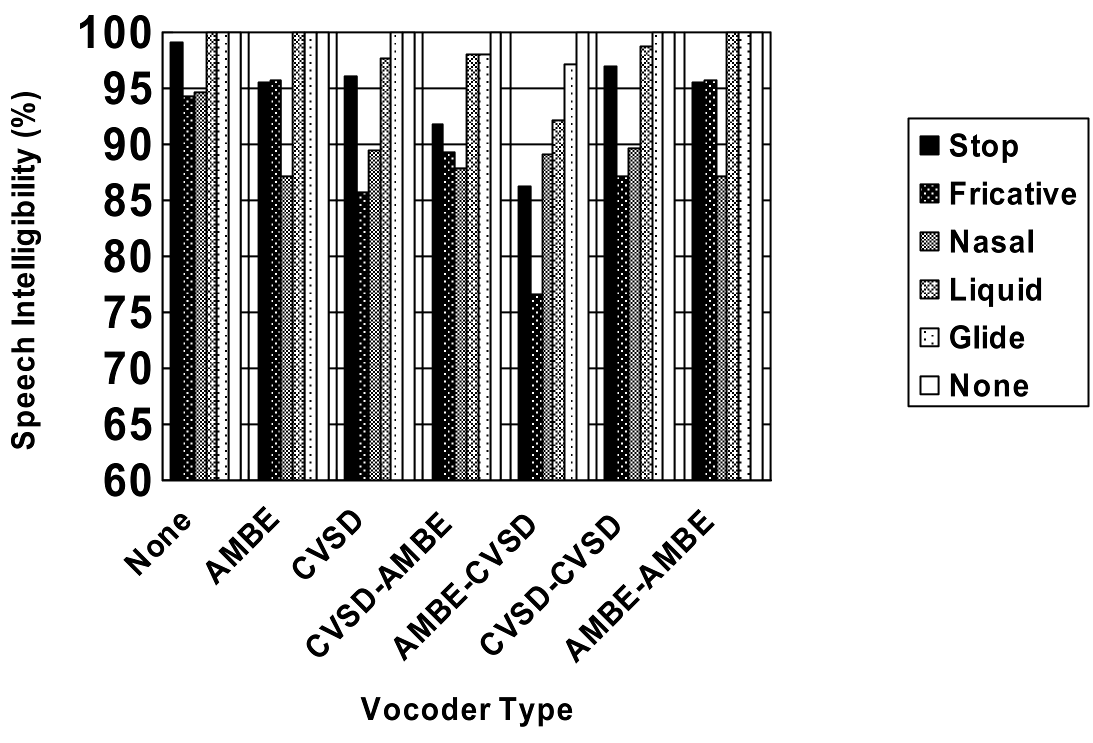

Figure 3: Speech intelligibility versus vocoder type and speech attribute with unaltered speech

A one-way analysis of variance (ANOVA) was performed on the speech intelligibility data. Percent correct scores were subjected to an arc-sine transformation before the ANOVA was performed [11]. The ANOVA revealed a main effect for vocoder type (F=2.75(6, 6), p=.042). A post hoc least significant difference (LSD) analysis was performed on the main effects to look for differences among vocoders. The control condition of no vocoding and the AMBE-to-CVSD tandem condition were found to be different than other vocoder conditions.

The data were further analyzed by the effects of vocoding on the manner of articulation. The speech attributes included stop, fricative, nasal, liquid, glide, and phonemic absence. The mean values are shown in Figure 3 for each of the seven vocoding conditions. A two-way ANOVA again revealed a main effect for vocoder type (F=12.75(6, 6), p=.006) and a main effect for attribute (F=14.53(6, 6), p=.002). A post hoc LSD analysis was performed on the speech attribute data. Stops, fricatives, and nasals were found not to be significantly different from each other. Liquid, glide, and absent attributes were also found not to be different from each other.

3.2 Clipped Speech Vocoded in Tandem

The second experiment was designed to measure the performance of the vocoders with hard clipped speech signals at 10 dB down from the peak. The mean and standard deviations for percent correct intelligibility are plotted in Figure 4. The percent correct values of speech intelligibility were measured to be 96.3% for the control condition of no vocoders, 92% for AMBE-alone, 87% for CVSD-alone, 84% for CVSD-AMBE, and 75% for AMBE-CVSD. In the like-vocoder condition, speech intelligibility was found to be 90.2% for CVSD-to-CVSD and 82.6% for AMBE-to-AMBE.

A one-way analysis of variance was performed on the speech intelligibility data. Percent correct scores were subjected to an arc-sine transformation before the ANOVA was performed [11]. The ANOVA revealed a main effect for vocoder type (F=2.45(6, 6), p=.048MP-HFM-123-11). A post hoc LSD analysis was performed on the data to look for differences among vocoders. The control condition of no vocoding, AMBE, and CVSD were found to be not different from each other. The four tandem conditions, AMBE-to-CVSD, CVSD-to-AMBE, AMBE-to-AMBE, and CVSD-to-CVSD, were found to be not different from each other.

The data were further analyzed by the effects of vocoding on the manner of articulation. The speech attributes included stop, fricative, nasal, liquid, glide, and phonemic absence. The mean values are shown in Figure 5 for each of the seven vocoding conditions. A two-way ANOVA again revealed a main effect for vocoder type (F=8.45(6, 6), p=.012) and a main effect for attribute (F=3.53(6.6), p=.042). A post hoc LSD analysis was performed on the speech attribute data. Stops, fricatives, and nasals were not found to be significantly different from each other. Liquid, glide, and absent attributes were also found not to be different from each other.

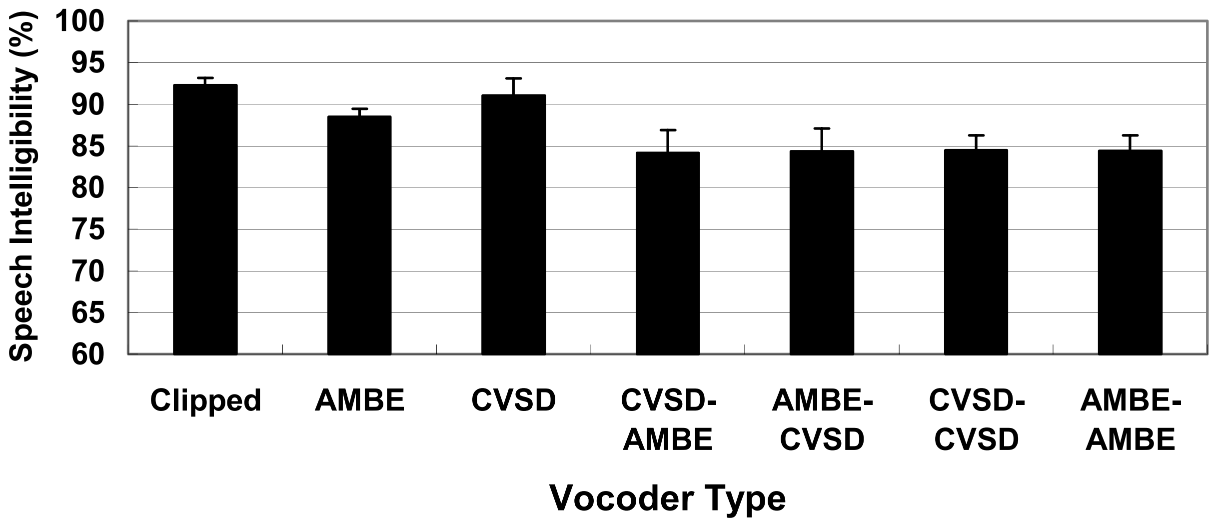

Figure 4: Speech intelligibility versus vocoder type with input speech signal hard clipped by 10dB

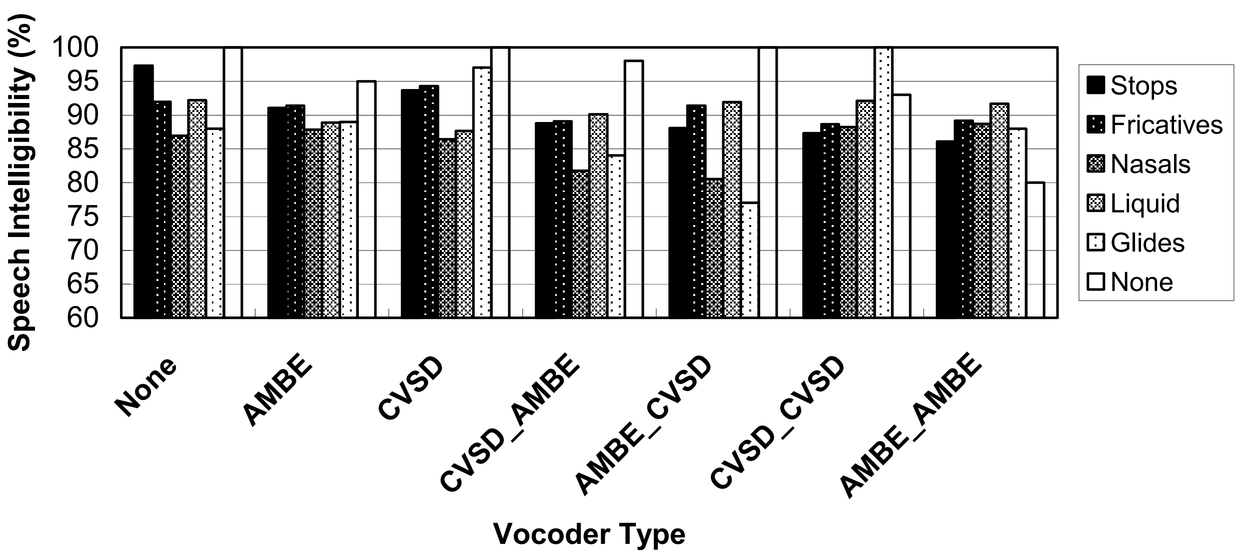

Figure 5: Speech intelligibility versus vocoder type with the speech signal hard clipped by 10dB

4.0 DISCUSSION

Both vocoders performed well with clean speech inputs. However, the AMBE parametric type of encoder performed much worse with the clipped speech than did the CVSD encoder. The CVSD vocoder was not able to process the output of the AMBE vocoder well. Conversely, the AMBE vocoder was able to process the output of the CVSD vocoder when unaltered speech was used. The CVSD vocoder performed well despite its simple algorithm and lack of knowledge of speech characteristics.

The effects of tandem vocoding and speech clipping were investigated in the current study. Many natural environmental factors and hostile jamming devices can potentially degrade the end-to-end speech intelligibility of the communication system. Deleterious effects include radio frequency channel noise, signal loss over long transmission distances, ambient acoustic noise at the talker or listener locations, encryption errors, bandwidth limitations, channel bit errors in the digital link, and burst disturbances. All of these factors and possibly more can have a degrading effect on speech intelligibility.

An implicit question in all speech intelligibility measurements is how much intelligibility is good enough for a given application. MIL-STD-1472F [12], the Department of Defense Design Criteria Standard for Human Engineering, suggests 91% intelligibility performance should be achieved on the MRT for operational military communication equipment and 97% performance for critical communications. Most military operators consider a voice communication system with 80% or better performance on the MRT to be acceptable, those between 70 and 80% to be marginal, and those below 70% to be unacceptable [4]. A further consideration is that air traffic controllers who are non-native speakers of English need a voice communication system with higher than normally acceptable speech intelligibility levels, i.e. greater than 80% with the MRT.

Since CVSD was developed in the 1970's, other waveform coding schemes have been created. Two such methods are adaptive differential pulse code modulation (ADPCM) and voice over internet protocol (VOIP). Each of these methods can out-perform CVSD in unperturbed, laboratory environments. However, all three react differently to disturbances in real-world environments. ADPCM is slightly more tolerant of bit error rates than CVSD, but requires more bandwidth to accomplish that level of performance [2]. VOIP issues deal with delay, jitter, and packet loss during transmission over digital communication links [6]. Further research is needed to understand the trade-offs between disturbances on speech intelligibility with VOIP techniques. The interaction of new and legacy vocoders when coupled in tandem should be investigated before such new systems are introduced into military communication systems.

5.0 CONCLUSIONS

Speech intelligibility performance of two types of vocoders was measured with clear, recorded speech and altered speech, hard clipped at 10 dB. The waveform vocoder was a continuous variable slope detector (CVSD) and the parametric encoder was the advanced multi-band excitation (AMBE) encoder. Intelligibility was measured in a control condition, with each type, in tandem with the opposite type, and in tandem with the same type. The AMBE and CVSD methods performed well in isolation, measured at 92.6 and 90.4% intelligibility, respectively. The CVSD-to-AMBE direction had little effect on intelligibility, measured at 89.2%. However, the AMBE-to-CVSD direction had a large degrading effect on intelligibility, 81.7%. The addition of a second vocoder of the same type had little effect on intelligibility. Intelligibility with the clipped speech input signals was generally lower than with the unaltered speech signals. Intelligibility levels in the tandem conditions with clipped speech were lower than with the unaltered speech, except for CVSD-AMBE.

A finer analysis of the effects of clipping and vocoding on speech intelligibility was evidenced in the phonemic analysis. Stops, fricatives, and nasals were adversely affected by the processing through a single vocoder and especially with dissimilar vocoders in tandem. Clipping had a large effect on all phoneme identifications. Liquids and glides were unaffected by vocoding in the clear speech condition, but were reduced in intelligibility by vocoding in the clipped speech condition.

The combination of two vocoders in tandem generally reduced performance in a non-linear way, especially in the clipped speech conditions. The asymmetry between waveform-to-parametric and parametric-to-waveform

encoders underscores the non-linear nature of tandem vocoders on intelligibility. Consideration should be given to the application of new vocoding techniques when embedded in military environments with legacy vocoders, such as the CVSD algorithm. Alternative methods of encoding speech signals are being explored to improve speech intelligibility performance in military communication systems.

6.0 REFERENCES

[1] McKinley, R. L. and Moore, T. J. (1985). *The effect of audio bandwidth on selected digital speech coding algorithms,* 1985 IEEE Military Communications Conference MILCOM'85, Boston, MA, October 20-23, 1985161-165, 9.1.1-9.1.5.

[2] McKinley, R. L. and Kurtz, R. V. (1984). *Subjective Performance of Selected Speech Coders in the Presence of Channel Errors* 1984 IEEE Military Communications Conference MILCOM'85, Los Angeles, California, October 21-24, 1985161-165, 9.1.1-9.1.5.

[3] http://www.dvsinc.com/papers/mbe.htm

[4] Moore, Thomas J. (1989). Speech Technology in the Cockpit, *Proceedings of Aviation Psych.*, 50-65.

[5] http://www.arcon.com/ddvpc/tandem.htm

[6] Atal, Bishnu S. (2005). 50 Years of Progress in Speech Waveform Coding, *Echoes The Newsletter of the Acoustical Society of America*, Volume 15(1), pp 1 and 4.

[7] Street, Michael D. (2002). Host Laboratory Role in the Selection of the Future NATO Narrow Band Voice Coder, IEEE Workshop on Speech Coding, L3.4, Tsukuba, Ibaraki, Japan, Oct 6-9, 2002.

[8] Military Communications, RTO MP-065, IST-023, February, 2002.

[9] McKinley, R. L. (1980). Voice Communication Research and Evaluation, AMRL-TR-80-25.

[10] ANSI Standard S3.2-1989 (rev. 1999) Methods for Measuring the Intelligibility of Speech Over Communication Systems.

[11] Studebaker, G. A. (1985). A "Rationalized" arcsine transformation. *Journal of Speech and Hearing Research* 28:455-462.

[12] Mil-Std-1472F (23 August 1999) Department of Defence Design Criteria Standard Human Engineering.